

**EV251222086**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

**APPLICATION FOR LETTERS PATENT**

**Audio Watermarking with Dual Watermarks**

Inventor(s):

Darko Kirovski

Henrique Malvar

Mariusz Jakubowski

ATTORNEY'S DOCKET NO. MS1-356US

1 **TECHNICAL FIELD**

2 This invention relates to systems and methods for protecting audio content.  
3 More particularly, this invention relates to watermarking audio data streams with  
4 two different watermarks.  
5

6 **BACKGROUND**

7 Music is the world's universal form of communication, touching every  
8 person of every culture on the globe. Behind the melody is a growing multi-  
9 billion dollar per year industry. This industry, however, is constantly plagued by  
10 lost revenues due to music piracy.

11 Piracy is not a new problem. But, as technologies change and improve,  
12 there are new challenges to protecting music content from illicit copying and theft.  
13 For instance, more producers are beginning to use the Internet to distribute music  
14 content. In this form of distribution, the content merely exists as a bit stream  
15 which, if left unprotected, can be easily copied and reproduced. At the end of  
16 1997, the International Federation of the Phonographic Industry (IFPI), the British  
17 Phonographic Industry, and the Recording Industry Association of America  
18 (RIAA) engaged in a project to survey the extent of unauthorized use of music on  
19 the Internet. The initial search indicated that at any one time there could be up to  
20 80,000 infringing MP3 files on the Internet. The actual number of servers on the  
21 Internet hosting infringing files was estimated to 2,000 with locations in over 30  
22 countries around the world.

23 Consequently, techniques for identifying copyright of digital audio content  
24 and in particular audio watermarking have received a great deal of attention in  
25 both the industrial community and the academic environment. One of the most

1 promising audio watermarking techniques is augmentation of a copyright  
2 watermark into the audio signal itself by altering the signal's frequency spectrum  
3 such that the perceptual characteristics of the original recording are preserved.  
4 The copy detection process is performed by synchronously correlating the  
5 suspected audio clip with the watermark of the content publisher. A common  
6 pitfall for all watermarking systems that facilitate this type of data hiding is  
7 intolerance to desynchronization attacks (e.g., sample cropping, insertion, and  
8 repetition, variable pitch-scale and time-scale modifications, audio restoration,  
9 combinations of different attacks) and deficiency of adequate techniques to  
10 address this problem during the detection process.

11 The business model of companies that deliver products for audio copyright  
12 enforcement has been focused on satisfying the minimal set of requirements in the  
13 IFPI's and RIAA's Request for Proposals (MUSE project) for technologies that  
14 inaudibly embed data in sound recordings. More recently, the RIAA has started  
15 the Secure Digital Music Initiative (SDMI) Forum in order to establish a standard  
16 for managing audio content copyrights. The requirements in both requests do not  
17 reflect accurately the common de-synch such as.

18 The existing techniques for watermarking discrete audio signals facilitate  
19 the insensitivity of the human auditory system (HAS) to certain audio phenomena.  
20 It has been demonstrated that, in the temporal domain, the HAS is insensitive to  
21 small signal level changes and peaks in the pre-echo and the decaying echo  
22 spectrum. The techniques developed to facilitate the first phenomenon are  
23 typically not resilient to de-synch attacks. Due to the difficulty of the echo  
24 cancellation problem, techniques which employ multiple decaying echoes to place  
25

1 a peak in the signal's cepstrum can hardly be attacked in real-time, but fairly easy  
2 using an off-line exhaustive search.

3 Watermarking techniques that embed secret data in the frequency domain  
4 of a signal facilitate the insensitivity of the HAS to small magnitude and phase  
5 changes. In both cases, publisher's secret key is encoded as a pseudo-random  
6 sequence that is used to guide the modification of each magnitude or phase  
7 component of the frequency domain. The modifications are performed either  
8 directly or shaped according to signal's envelope. In addition, a watermarking  
9 scheme has been developed which facilitates the advantages but also suffers from  
10 the disadvantages of hiding data in both the time and frequency domain. All  
11 reported approaches perform the watermark detection process on both the audible  
12 and inaudible spectrum components, thus enabling the attacker to reduce the  
13 correlation between the watermarked signal and its watermark by adding noise in  
14 the inaudible domain. Similarly, it has not been demonstrated whether these  
15 watermarking schemes would survive combinations of common attacks: de-synch  
16 in both the temporal and frequency domain and mosaic-like attacks.

17 Accordingly, there is a need for a new framework of protocols for hiding  
18 and detecting watermarks in digital audio signals that are effective against  
19 desynchronization attacks. The framework should possess several attributes,  
20 including perceptual invisibility (i.e., the embedded information should not induce  
21 audible changes in the audio quality of the resulting watermarked signal) and  
22 statistical invisibility (i.e., the embedded information should be quantitatively  
23 imperceptible for any exhaustive, heuristic, or probabilistic attempt to detect or  
24 remove the watermark). Additionally, the framework should be tamperproof (i.e.,  
25 an attempt to remove the watermark should damage the value of the music well

1 above the hearing threshold) and inexpensive to license and implement on both  
2 programmable and application-specific platforms. The framework should be such  
3 that the process of proving audio content copyright both in-situ and in-court does  
4 not involve usage of the original recording.

5 The framework should also be flexible to enable a spectrum of protection  
6 levels, which correspond to variable audio presentation and compression  
7 standards, and yet resilient to common attacks spawned by powerful digital sound  
8 editing tools. The standard set of plausible attacks is itemized in the IFPI's and  
9 RIAA's Request for Proposals and, among others, it encapsulates the following  
10 security requirements:

- 11 • Two successive D/A and A/D conversions;
  - 12 • Data reduction coding techniques such as MP3;
  - 13 • Adaptive transform coding;
  - 14 • Adaptive subband coding;
  - 15 • Digital Audio Broadcasting (DAB);
  - 16 • Dolby AC2 and AC3 systems;
  - 17 • Applying additive or multiplicative noise;
  - 18 • Applying a second Embedded Signal, using the same system, to a single  
19 program fragment;
  - 20 • Frequency response distortion corresponding to normal analogue  
21 frequency response controls such as bass, mid and treble controls, with  
22 maximum variation of 15 dB with respect to the original signal; and
  - 23 • Applying frequency notches with possible frequency hopping.
- 24  
25

## SUMMARY

This invention concerns an audio watermarking technology for inserting and detecting strong and weak watermarks in audio signals. The strong watermark identifies the content producer, providing a signature that is embedded in the audio signal and cannot be removed. The strong watermark is designed to survive all typical kinds of processing, including compression, equalization, D/A and A/D conversion, recording on analog tape, and so forth. It is also designed to survive malicious attacks that attempt to remove the watermark from the signal, including changes in time and frequency scales, pitch shifting, and cut/paste editing.

The weak watermark identifies the content as an original. With the exception of D/A and A/D conversion with good fidelity, other kinds of processing (especially compression) significantly remove the weak watermark. In this manner, an audio signal can be readily identified as an original or a copy depending upon the presence or absence of the weak watermark signature.

In one described implementation, a watermark encoding system is implemented at a content provider/producer to encode the audio signal with both a strong and a weak watermark. The watermark encoding system has a converter to convert an audio signal into frequency and phase components and a mask processor to determine a hearing threshold for corresponding frequency components. The watermark encoding system also has a pattern generator to generate both the strong and weak watermarks, and a watermark insertion unit to selectively insert either the strong or weak watermark into the audio signal. More particularly, the watermark insertion unit adds the strong watermark to the audio signal when the signal exceeds the hearing threshold by a buffer value (e.g., 1-8 dB). If the signal falls below the hearing threshold by more than the buffer value,

1 the watermark insertion unit adds the weak watermark component to the audio  
2 signal. When the signal falls within the buffer area about the hearing threshold,  
3 the insertion unit takes no action because the signal component is not significantly  
4 above or below the threshold to be watermarked.

5 A watermark detecting system is implemented at a client that plays the  
6 audio clip. Like the encoding system, the watermark detecting system has the  
7 converter, the mask processor, and the watermark pattern generator. It is also  
8 equipped with a watermark detector that locates any strong and weak watermarks  
9 in the audio clip. The watermark detector determines which block interval of the  
10 watermarked audio signal contains the watermark pattern and if the strong or weak  
11 watermark generated by a particular set of keys is present in that block interval of  
12 the signal.

### 13 14 **BRIEF DESCRIPTION OF THE DRAWINGS**

15 The same numbers are used throughout the drawings to reference like  
16 elements and features.

17 Fig. 1 is a block diagram of an audio production and distribution system in  
18 which a content producer/provider watermarks audio signals and subsequently  
19 distributes that watermarked audio stream to a client over a network.

20 Fig. 2 is a block diagram of a watermarking encoding unit implemented, for  
21 example, at the content producer/provider.

22 Fig. 3 is a frequency domain representation of an audio signal along with  
23 corresponding strong and weak watermarking components.

24 Fig. 4 is a flow diagram showing the watermarking process of inserting  
25 strong and weak watermarks into an audio signal.

1        Fig. 5 is a block diagram of a watermarking detecting unit implemented, for  
2 example, at the client.

3        Fig. 6 is a flow diagram showing a watermark detection process of  
4 detecting strong and weak watermarks in an audio signal.

5        Fig. 7 show time-scale plots of normalized correlation values used to detect  
6 presence and absence of a watermark.

7        Fig. 8 shows plots of the distribution of normalized correlation for four  
8 different artists.

9        Fig. 9 is a block diagram of a watermarking encoding unit implemented  
10 according to a second implementation.

11       Fig. 10 is a block diagram of a watermarking detecting unit implemented  
12 according to a second implementation.

### 13 14 **DETAILED DESCRIPTION**

15       Fig. 1 shows an audio production and distribution system 20 having a  
16 content producer/provider 22 that produces original musical content and  
17 distributes the musical content over a network 24 to a client 26. The content  
18 producer/provider 22 has a content storage 30 to store digital audio streams of  
19 original musical content. The content producer 22 has a watermark encoding  
20 system 32 to sign the audio data stream with a watermark that uniquely identifies  
21 the content as original. The watermark encoding system 32 may be implemented  
22 as a standalone process or incorporated into other applications or an operating  
23 system.

24       A watermark is an array of bits generated using a cryptographically secure  
25 pseudo-random bit generator and a new error correction encoder. The pseudo-



1 uniqueness of each watermark is provided by initiating the bit generator with a key  
2 unique to each audio content publisher. The watermark is embedded into a digital  
3 audio signal by altering its frequency magnitudes such that the perceptual audio  
4 characteristics of the original recording are preserved. Each magnitude in the  
5 frequency spectrum is altered according to the appropriate bit in the watermark.

6 The watermark encoding system 32 applies two types of watermarks: a  
7 strong watermark and a weak watermark. The strong watermark identifies the  
8 content producer 22, providing a signature that is embedded in the audio signal  
9 and cannot be removed. The strong watermark is designed to survive all typical  
10 kinds of processing, including compression, equalization, D/A and A/D  
11 conversion, recording on analog tape, and so forth. It is also designed to survive  
12 malicious attacks that attempt to remove the watermark from the signal, including  
13 changes in time and frequency scales, pitch shifting, and cut/paste editing. The  
14 weak watermark identifies the content as an original. With the exception of D/A  
15 and A/D conversion with good fidelity, other kinds of processing (especially  
16 compression) significantly remove the weak watermark. In this manner, an audio  
17 signal can be readily identified as an original or a copy depending upon the  
18 presence or absence of the weak watermark signature.

19 The content producer/provider 22 has a distribution server 34 that streams  
20 the watermarked audio content over the network 24 (e.g., Internet). An audio  
21 stream with both watermarks embedded therein represents to a recipient that the  
22 stream is original and being distributed in accordance with the copyright authority  
23 of the content producer/provider 22. The server 34 may further compress and/or  
24 encrypt the content conventional compression and encryption techniques prior to  
25 distributing the content over the network 24.

1 The client 26 is equipped with a processor 40, a memory 42, and one or  
2 more media output devices 44. The processor 40 runs various tools to process the  
3 audio stream, such as tools to decompress the stream, decrypt the data, filter the  
4 content, and/or apply audio controls (tone, volume, etc.). The memory 42 stores  
5 an operating system 50, such as a Windows brand operating system from  
6 Microsoft Corporation, which executes on the processor. The client 26 may be  
7 embodied in a many different ways, including a computer, a handheld  
8 entertainment device, a set-top box, a television, an audio appliance, and so forth.

9 The operating system 50 implements a client-side watermark detecting  
10 system 52 to detect the strong and weak watermarks in the audio stream and a  
11 media audio player 54 to facilitate play of the audio content through the media  
12 output device(s) 44 (e.g., sound card, speakers, etc.). If both watermarks are  
13 present, the client is assured that the content is original and can be played.  
14 Absence of the weak watermark indicates that the audio stream is a copy of an  
15 original. If both watermarks are absent, the content is neither a protected original  
16 nor a copy of a protected original. The operating system 50 and/or processor 40  
17 may be configured to enforce certain rules imposed by the content  
18 producer/provider (or copyright owner). For instance, the operating system and/or  
19 processor may be configured to reject fake or copied content that does not possess  
20 both strong and weak watermarks. In another example, the system could play  
21 unverified content with a reduced level of fidelity.

### 22 23 **Dual Watermark Insertion**

24 Fig. 2 shows one implementation of the watermark encoding system 32. It  
25 receives an original audio signal  $x(n)$  and produces a watermarked audio signal

1  $y(n)$ . The original signal is processed in blocks of  $M$  samples and stored in the  
2 content storage 30 (Fig. 1). Typically,  $M$  is set to 2,048 for CD-quality signals  
3 sampled at 44.1 kHz, corresponding to a block time duration of about 46.4 ms.  
4 The encoding system 32 has an MCLT (modulated complex lapped transform)  
5 component 60 that transforms the input signal  $x(n)$  to the frequency domain,  
6 producing the vector  $X(k)$  also with  $M$  components (i.e.,  $k = 0, 1, \dots, M-1$ ). Each  
7  $X(k)$  is a complex number, and  $X_{MAG}(k)$  is referred to as its magnitude and  $\phi(k)$  as  
8 its phase. The magnitude is measured in a logarithmic scale, in decibels (dB).  
9 One specific implementation of the MCLT component 60 is described in more  
10 detail in a co-pending patent application, entitled "A system and Method for  
11 Producing Modulated Complex Lapped Transforms", which was filed February  
12 26, 1999 and is assigned to Microsoft Corporation. This application is  
13 incorporated by reference.

14 The magnitude frequency components  $X_{MAG}(k)$  are processed by an auditory  
15 masking model processor 62, which computes a set of hearing thresholds  $z(k)$  ( $k =$   
16  $0, 1, \dots, M-1$ ), one for each frequency. The auditory masking model processor 62  
17 simulates the dynamics of the human ear and computes  $z(k)$  such that  $X_{MAG}(k)$  is  
18 audible only if its value is above  $z(k)$ . One example implementation of a masking  
19 model is a codec employed in "MSAudio", a product available from Microsoft  
20 Corporation. This codec is described in a co-pending U.S. patent application  
21 serial number 09/085,620, entitled "Scalable Audio Coder and Decoder", which  
22 was filed May 27, 1998 and is assigned to Microsoft Corporation. This  
23 application is incorporated by reference.

24 Fig. 3 is a frequency domain plot 90 showing samples of the audio signal's  
25 magnitude frequency components  $X_{MAG}(k)$ . The auditory masking model

1 processor 62 computes a hearing threshold from the magnitude frequency  
2 components that dictate whether an individual sample is audible or not. In this  
3 illustration, samples rising above the threshold are audible, whereas samples  
4 falling below the threshold are not audible.

5 With reference again to Fig. 2, a pattern generator 64 creates strong and  
6 weak watermark signatures that will be selectively mixed with the audio signal.  
7 The pattern generator is illustrated as having a strong watermark generator 66 to  
8 produce a strong watermark vector  $w(k)$  using a cryptographic algorithm  
9 controlled by a key  $K_S$ . The pattern generator 64 also has a weak watermark  
10 generator 68 to produce a weak watermark vector  $u(k)$  using a cryptographic  
11 algorithm controlled by a key  $K_W$ . The strong and weak generators 66 and 68 may  
12 be implemented separately, or integrated as the same unit with the only difference  
13 being the key used to produce the desired strong or weak pattern.

14 A new vector is only generated for every  $L$  blocks, which constitute a  
15 frame. The parameter  $L$  is typically set to 10, as discussed below. Also, the strong  
16 watermark vector  $w(k)$  is such that  $w(k)$  remains constant for a group of  
17 frequencies, e.g.  $w(0) = w(1) = \dots = w(N_0)$ ,  $w(N_0+1) = w(N_0+2) = \dots = w(N_1)$ , and  
18 so forth, with the parameters  $N_0$ ,  $N_1$ , etc. typically approximating a Bark frequency  
19 scale or another appropriate frequency scale.

20 The components of the strong watermark vector  $w(k)$  and weak watermark  
21 vector  $u(k)$  are binary entries, with values equal to  $-Q$  or  $+Q$  (in decibels). In a  
22 typical application,  $Q$  may be set to 1 dB, for example. The keys and  
23 cryptographic algorithm are selected such that the strong and weak watermark  
24 values have zero mean, meaning that any given value is equally likely to assume  
25 values  $+Q$  or  $-Q$ .

Fig. 3 shows frequency plot 92 with a few samples from the strong watermark vector and a frequency plot 94 with a few samples from the weak watermark vector  $u(k)$ . The patterns are generated based upon the respective strong and weak keys  $K_S$  and  $K_W$ .

The watermark encoding system 32 has a watermark insertion unit 70 that selectively combines either the strong watermark vector  $w(k)$  or the weak watermark vector  $u(k)$  with the magnitude frequency components  $X_{MAG}(k)$  from MCLT component 60 based upon the hearing threshold vector  $z(k)$  from masking model 62. The watermark insertion unit 70 has multiple insertion operators  $72(0)$ ,  $72(1), \dots, 72(k)$  ( $k = 0, 1, \dots, M-1$ ) for each corresponding frequency. In this manner, for each frequency index  $k$ , the magnitude frequency components  $X_{MAG}(k)$  is modified to generate the watermarked magnitude frequency components  $Y_{MAG}(k)$ . More specifically, each insertion operation modifies its magnitude frequency components  $X_{MAG}(k)$  with the strong watermark value  $w(k)$  if the magnitude frequency component exceeds the hearing threshold  $z(k)$  and alternatively, with the weak watermark value  $u(k)$  if the magnitude frequency component fails to exceed the hearing threshold  $z(k)$ . The insertion process is described below in more detail with reference to Figs. 3 and 4.

An IMCLT (Inverse MCLT) component 80 receives the watermarked magnitude frequency components  $Y_{MAG}(k)$  from the watermark insertion unit 70 and the phases  $\phi(k)$  from the MCLT component 60. The IMCLT component 80 converts the frequency-domain signal  $\{Y_{MAG}(k), \phi(k)\}$  to a time-domain watermarked signal block  $y(n)$ . The time domain audio signal is in a form that can then be stored in the content storage 30 and/or distributed over the network 24 to the client 26.

1       The insertion process is repeated through a group of  $T$  blocks. The  
2 parameter  $T$  controls the length of the watermark, and is typically set between 20  
3 and 300 blocks. Larger values of  $T$  result in more reliable detection, as described  
4 below.

5       Fig. 4 shows a watermark insertion process performed by the watermark  
6 insertion unit 70. These steps may be performed in software, hardware, or a  
7 combination thereof. At the start of the process, the watermark insertion unit 70  
8 reads the magnitude frequency components  $X_{MAG}(k)$ , the hearing thresholds  $z(k)$ ,  
9 the strong watermark vector  $w(k)$ , and the weak watermark vector  $u(k)$  (steps 100  
10 and 102). Corresponding values in these vectors are passed to respective insertion  
11 operators 72(0)-72( $M$ -1). After the frequency is initialized (i.e.,  $k=0$ ) (step 104),  
12 the watermark insertion unit 70 begins cycling through the  $M$  samples and  
13 determining whether any given signal rises above an associated hearing threshold,  
14 resulting in application of a strong watermark, or falls below the hearing threshold,  
15 resulting in application of the weak watermark.

16       At step 106, the  $k^{th}$  insertion operator 72( $k$ ) evaluates whether the  
17 magnitude frequency components  $X_{MAG}(k)$  is greater than the hearing threshold  
18  $z(k)$  plus a buffer value  $B$ . If it is, the insertion operator 72( $k$ ) adds the strong  
19 watermark component  $w(k)$  to the magnitude frequency components  $X_{MAG}(k)$  to  
20 produce the watermarked magnitude frequency component  $Y(k)$  (step 108).  
21 Referring to Fig. 3, sample 96a is an example of the situation where the signal  
22 exceeds the hearing threshold by a value  $B$  (not shown), and hence this sample  
23 would be reduced by  $-Q$  as a result of the associated watermark component 96b.

24       If the signal does not exceed the hearing threshold by a value  $B$ , the  
25 insertion operator 72( $k$ ) discerns whether the magnitude frequency components

1  $X_{MAG}(k)$  is less than the hearing threshold  $z(k)$  minus a buffer value  $B$  (step 110) If  
2 so, the insertion operator 72(k) adds the weak watermark component  $u(k)$  to the  
3 magnitude frequency components  $X_{MAG}(k)$  to produce the watermarked magnitude  
4 frequency component  $Y(k)$  (step 112). Referring to Fig. 3, sample 98a is an  
5 example of the situation where the signal falls below the hearing threshold by a  
6 value  $B$  (not shown), and hence this sample is increased by  $Q$  as a result of the  
7 associated watermark component 98b.

8 If the signal fails to exceed or be less than the hearing threshold by a value  
9  $B$ , the insertion operator takes no action. The buffer value  $B$  thus defines a dead  
10 zone about the threshold region for which the signal component is not significantly  
11 above or below the threshold to be watermarked. Typical values of  $B$  range from  
12 1 dB to 8 dB.

13 At step 114, the watermark insertion unit 70 proceeds to the next frequency  
14 (i.e.,  $k=k+1$ ). Assuming this is not the last  $M$  sample (i.e., step 116), the dual  
15 watermark analysis continues for the next signal sample. However, once the  
16 watermark insertion unit 70 processes all  $M$  samples, it writes the watermarked  
17 vector  $Y(k)$  to the IMCLT component 80 and the process is completed for this  
18 block (steps 118 and 120).

19 This insertion process advantageously provides two different watermarks  
20 with different purposes. The strong watermark is firmly embedded into the  
21 audible signal. The strong watermark cannot be removed and survives all typical  
22 kinds of processing as well as malicious attacks that attempt to remove the  
23 watermark from the signal. The weak watermark is lightly implanted into the non-  
24 audible portions of the signal. These are the samples most likely to be removed  
25

1 during signal processing (e.g., compression) and hence provide a valuable  
2 indication as to whether the audio signal is a copy, rather than an original.

### 4 Watermark Detection

5 Fig. 5 shows one implementation of the watermark decoding system 52 that  
6 executes on the client 26 to detect whether the content is original or a copy (or  
7 fake). To detect the strong and weak watermarks, the system finds whether the  
8 corresponding patterns  $\{w(k)\}$  and  $\{u(k)\}$  are present in the signal.

9 Like the encoder system 32, the watermark decoding system 52 has an  
10 MCLT component 60, an auditory masking model 62, and a pattern generator 64.  
11 The MCLT component 60 receives a decoded audio signal  $y(n)$  and transforms the  
12 signal to the frequency domain, producing the vector  $Y(k)$  having a magnitude  
13 component  $Y_{MAG}(k)$  and phase component  $\phi(k)$ . The auditory masking model 62  
14 computes a set of hearing thresholds  $z(k)$  ( $k = 0, 1, \dots, M-1$ ) based on the  
15 magnitude components  $Y_{MAG}(k)$ . Since the thresholds are computed from  $Y_{MAG}(k)$ ,  
16 as opposed to  $X_{MAG}(k)$ , the threshold vector  $z(k)$  will not be identical to the vector  
17  $z(k)$  computed at the insertion unit 70, but the small differences caused by the  
18 watermarks do not affect operation of the watermark detector. A pattern generator  
19 64 creates strong and weak watermark vectors  $w(k)$  and  $u(k)$ .

20 Unlike the encoder system 32, the watermarking decoding system 52 has a  
21 watermark detector 130 that processes all available blocks of the watermarked  
22 signal  $\{Y_{MAG}(k)\}$ , the hearing thresholds  $\{z(k)\}$ , and the strong and weak  
23 watermark patterns  $\{w(k)\}$  and  $\{u(k)\}$ . The watermark detector 130 has a  
24 synchronization searcher 132, a correlation peak seeker 134, and a random  
25 operator 136. The decoding system 52 also has a random number generator



(RNG) 140 that provides a random variable  $\varepsilon$  to the watermark detector 130 to thwart a sample-by-sample attack. The operation of these modules is described below in more detail with reference to Fig. 6.

In general, there are two basic problems in detecting the watermark patterns:

1. Determine which  $T$ -block interval of the watermarked audio signal contains the watermark pattern. This is the synchronization problem.
2. Detect if the watermark corresponding to a particular set of keys  $K_S$  and  $K_W$  is present in that  $T$ -block interval of the signal.

The two problems are related and are solved in conjunction. So, for discussion purposes, assume that there is perfect synchronization in that the location of the  $T$ -block watermark interval is known. This removes the first problem, which will be addressed below in more detail. Also, assume that the detection process is focused on detecting only the strong watermark. The process for detecting the weak watermark is the same, except that the weak watermark pattern  $\{u(k)\}$  replaces the strong watermark pattern  $\{w(k)\}$ .

Let  $\mathbf{y}$  be a vector formed by all coefficients  $\{Y(k)\}$ . Furthermore, let  $\mathbf{x}$ ,  $\mathbf{z}$ , and  $\mathbf{w}$  be vectors formed by all coefficients  $\{X(k)\}$ ,  $\{z(k)\}$ , and  $\{w(k)\}$ , respectively. All values are in decibels (i.e., in a log scale). Furthermore, let  $y(i)$  be the  $i^{\text{th}}$  element of a vector  $\mathbf{y}$ . The index  $i$  varies from 0 to  $K-1$ , where  $K = TM$ .

Watermark insertion is given by,

$$\mathbf{y} = \mathbf{x} + \mathbf{w}, \text{ or } y(i) = x(i) + w(i), i = 0, 1, \dots, K-1 \quad (1)$$

where the actual vector  $\mathbf{w}$  may have some of its elements set to zero, depending on the values of the hearing threshold vector  $\mathbf{z}$ . Note that strictly speaking the sum in Equation (1) is not a linear superposition, because the values  $w(i)$  are modified based on  $v(i)$ , which in turn depends on the signal components  $x(i)$ .

Now, consider a correlation operator  $NC$  defined as follows:

$$NC \equiv \frac{\sum_{i=0}^{K-1} y(i)w(i)}{\sum_{i=0}^{K-1} w^2(i)} \quad (2)$$

In the case where the signal is not watermarked,  $y(i) = x(i)$  and the correlation measure is equal to:

$$NC_0 \equiv \frac{\sum_{i=0}^{K-1} x(i)w(i)}{\sum_{i=0}^{K-1} w^2(i)} \quad (3)$$

Since the watermark values  $w(i)$  have zero mean, the numerator in Equation (3) will be a sum of negative and positive values, whereas the denominator will be equal to  $Q^2$  times the number of indices in the set  $\mathbf{I}$ . Therefore, for a large  $K$ , the

measure  $NC_0$  will be a random variable with an approximately normal (Gaussian) probability distribution, with an expected value of zero and a variance much smaller than one.

In the case where the signal is watermarked,  $y(i) = x(i) + w(i)$  and the correlation measure is equal to:

$$NC_1 \equiv \frac{\sum_{i=0}^{K-1} y(i)w(i)}{\sum_{i=0}^{K-1} w^2(i)} = \frac{\sum_{i=0}^{K-1} [x(i) + w(i)]w(i)}{\sum_{i=0}^{K-1} w^2(i)} = NC_0 + 1 \quad (4)$$

As seen in Equation (4), if the watermark is present, the correlation measure will be close to one. More precisely,  $NC_1$  will be a random variable with an approximately normal probability distribution, with an expected value of one and a variance much smaller than one.

The correlation peak seeker 134 in the watermark detector 130 determines the correlation operator  $NC$ . From the value of the correlation operator  $NC$ , the watermark detector 130 decides whether a watermark is present or absent. In its most basic form, the watermark presence decision compares the correlation operator  $NC$  to a detection threshold "Th", forming the following simple rule:

- If  $NC \leq Th$ , the watermark is not present.
- If  $NC > Th$ , the watermark is present.

1       The detection threshold “Th” is a parameter that controls the probabilities  
2 of the two kinds of errors:

- 3  
4       1. False alarm: the watermark is not present, but is detected as being  
5       present.
- 6       2. Miss: the watermark is present, but is detected as being absent.

7  
8       If  $Th = 0.5$ , the probability of a false alarm “Prob(false alarm)” equals the  
9 probability of a miss “Prob(miss)”. However, in practice, it is typically more  
10 desirable that the detection mechanism error on the side of never missing detection  
11 of a watermark, even if in some cases one is falsely detected. This means that  
12  $Prob(miss) \ll Prob(false\ alarm)$  and hence, the detection threshold is set to  
13  $Th < 0.5$ . In some applications false alarms may have a higher cost. For those, the  
14 detection threshold is set to  $Th > 0.5$ .

15       The decision rule may be slightly modified to account for a small random  
16 variance “ $\epsilon$ ” generated by the random number generator 140 (Fig. 5). The  
17 modified rule is as follows:

- 18  
19       • If  $NC < Th + \epsilon$ , the watermark is not present.
- 20       • If  $NC > Th + \epsilon$ , the watermark is present.

21  
22       The random threshold correction  $\epsilon$  is a random variable with a zero mean  
23 and a small variance (typically around 0.1 or less). It is preferably truly random  
24 (e.g. generated by reading noise values on a physical device, such as a zener  
25 diode).

1       The slightly randomized decision rule protects the system against attacks  
2 that modify the watermarked signal until the detector starts to fail. Such attacks  
3 could potentially learn the watermark pattern  $w(i)$  one element at a time, even if at  
4 a high computational cost. By adding the noise  $\varepsilon$  to the decision rule, such attacks  
5 are prevented from working.

6       Returning to the synchronization problem, the test watermark pattern and  
7 the watermarked signal need to be aligned for the correlation detector to work  
8 properly. This means that the strong watermark values  $w(i)$  (or weak watermark  
9 values  $u(i)$ ) in the test pattern and watermarked signal match. If not, the expected  
10 value of NC decays rapidly from one.

11       The synchronization searcher module 132 finds the right sync point by  
12 searching through a sequence of starting points for the  $T$ -block group of samples  
13 that will be used to build the signal vector. A sync point  $r$  is initialized (i.e.,  $r = 0$ )  
14 and incremented in steps  $R$ . At each interval, the correlation peak seeker module  
15 134 recomputes the correlation  $NC(r)$ . The true correlation is chosen as:

$$NC = \max_r NC(r) \quad (5)$$

16  
17  
18  
19  
20       The sync point increment  $R$  is set such that  $NC(r)$  and  $NC(r+R)$  differ  
21 significantly. If  $R$  is set to one, for example, an excessive amount of computations  
22 will be performed. In practice,  $R$  is typically set to about 10–50% of the block size  
23  $M$ .

24       Fig. 6 shows a watermark detection process performed by the watermark  
25 detector 130. These steps may be performed in software, hardware, or a

1 combination thereof. The process is illustrated as detecting the strong watermark  
2  $w(k)$ , but the weak watermark can be detected using the same process, replacing  
3 the strong watermark pattern  $\{w(i)\}$  with the weak watermark pattern  $\{u(i)\}$ .

4 At the start of the process, the watermark pattern generator 64 generates a  
5 strong watermark vector  $\{w(i)\}$  using the strong key  $K_S$  (steps 150 and 152). The  
6 detecting system 52 allocates buffer for a correlation array  $\{NC(r)\}$  that will be  
7 computed (step 154) and initializes the sync point  $r$  to a first sample (step 156).

8 At step 158, the MCLT module 60 reads in the audio signal  $y(n)$ , starting at  
9  $y(r)$ , and computes the magnitude values  $Y_{MAG}(k)$ . The auditory masking model 62  
10 then computes the hearing threshold  $z(k)$  from  $Y_{MAG}(k)$  (step 160). The strong  
11 watermark, magnitude frequency components, and hearing thresholds are passed  
12 to the watermark detector 130.

13 At step 162, the watermark detector 130 tests for a condition where there is  
14 no watermark by setting the watermark vector  $w(i)$  to zero, such that the  
15 watermarked input vector  $Y(i)$  is less than the hearing threshold by buffer value  $B$ .  
16 The watermark detector 130 then computes the correlation value  $NC$  for the sync  
17 point  $r$  (step 164). The process of computing correlation values  $NC$  continues for  
18 subsequent sync points, each incremented from the previous point by step  $R$  (i.e.,  $r$   
19  $= r + R$ ) (step 166), until correlation values for a maximum number of sync points  
20 has been collected (step 168).

21 At step 170, the watermark detector 130 reads the detection threshold "Th"  
22 and generates the random threshold correction  $\varepsilon$ . More particularly, the random  
23 operator 136 computes the random threshold correction  $\varepsilon$  based on a random  
24 output from the random number generator 140. Then, at step 172, the correlation  
25 peak seeker 134 searches for peak correlation such that:

$$NC = \max_r NC(r)$$

If the correlation value  $NC > Th + \epsilon$ , the watermark is present and a decision flag  $D$  is set to one (steps 174 and 176). Otherwise, the watermark is not present and the decision flag  $D$  is reset to zero (step 178). The watermark detector 130 writes the decision value  $D$  and the process concludes (steps 180 and 182).

The process in Fig. 6 is repeated or performed concurrently to detect whether the weak watermark is present. The only difference in the process for detecting the weak watermark is that the strong watermark pattern vector  $w(i)$  is replaced by the weak watermark pattern vector  $u(i)$ , and step 162 is modified to set  $u(i) = 0$  when  $Y(i)$  is higher than the hearing threshold by the buffer value  $B$ .

After the decision values have been computed for both the strong and weak watermarks, the watermark detector 130 outputs two flags. A strong watermark presence flag  $O_s$  indicates whether the strong watermark is present and a weak watermark presence flag  $O_w$  indicates whether the weak watermark is present. If both watermarks are present, the audio content is original. Absence of the weak watermark indicates that the audio stream is a copy of an original. If both watermarks are absent, the content is neither original nor a copy of an original.

Fig. 7 depicts time-scale plots of normalized correlation values obtained from the watermark detector 130 during a search for a watermark in an audio clip. Plots 184a and 184b demonstrate an audio clip that has been watermarked. A peak of values of the normalized correlation illustrated in plots 184a and 184b clearly indicates existence and location of the watermark. Plots 186a and 186b demonstrate an audio clip that has not been correlated with the test watermark.

1 A number of experiments were performed to determine the distributions of  
2 normalized correlation for different watermarking schemes. Each experiment was  
3 conducted on four representative audio samples (composers: Wolfgang Amadeus  
4 Mozart, Pat Metheney, Tracy Chapman, and Alanis Morissette). Each benchmark  
5 audio clip was watermarked 500 times. Correlation tests were performed for each  
6 watermarked version of the audio clip, one with a correct watermark and 99 with  
7 incorrect watermarks. There was no significant difference of statistical behavior of  
8 the applied watermarking scheme for any of the benchmark audio clips.

9 Fig. 8 depicts the results obtained from four different evaluations of the  
10 distribution of normalized correlation. Each row of diagrams in Fig. 8 depicts the  
11 results for one of the following four watermarking schemes:

- 12
- 13 (i)  $d_{\text{boffset}}=2\text{dB}$ ,  $\text{DFS}=1\%$ , fair cut of inaudible portion of frequency  
14 spectrum;
- 15 (ii)  $d_{\text{boffset}} = 2\text{dB}$ ,  $\text{DFS}=1\%$ , correlation test performed on the entire  
16 frequency spectrum;
- 17 (iii)  $d_{\text{boffset}} = 2\text{dB}$ ,  $\text{DFS}=0.5\%$ , fair cut of inaudible portion of the  
18 frequency spectrum; and
- 19 (iv)  $d_{\text{boffset}} = 2\text{dB}$ ,  $\text{DFS}=1\%$ , unfair cut of the inaudible portion of the  
20 frequency spectrum.
- 21

22 For each tested watermarking scheme, the following information is  
23 displayed in each column of the diagrams in Fig. 8:

24

25



1 a diagram of the convergence of a normalized correlation as well as  
2 the standard deviation of the distribution;

- 3 • a diagram that quantifies the probability of a false alarm; and
- 4 • a diagram that quantifies the probability of misdetection for a given  
5 length of the watermark sequence (X-axis on all diagrams).

6  
7 The depicted information clearly indicates that the consideration of only the  
8 audible portion of the audio clip as well as the fairness of its selection improves  
9 the confidence in making a decision for a particular value of the correlation for  
10 several orders of magnitude.

11 For further evaluation of the security of the content protection mechanism,  
12 we have selected a representative algorithm with the following properties:

- 13
- 14 • **Window size** = 4096 time-domain samples,
- 15 • **Number of bits embedded per window** = 153 bits,
- 16 • **Dynamic frequency shift (DFS)** =  $\pm 0.5\%$
- 17 • **Dynamic time warping (DTW)** =  $\pm 0.75\%$ ,
- 18 •  **$R$  - redundancy in time** = 20 windows,  **$M$**  = 10 windows,
- 19 •  **$L_{\text{MIN}}$**  = 45 ~ 45 seconds, **Decision Threshold  $\text{Th}$**  = 0.70,
- 20 •  **$P_{\text{FA}} < \Omega = 10^{-9}$** , and  **$P_{\text{MD}} < \Xi = 10^{-2}$** .
- 21

22 If it is assumed that the watermark is embedded into an audio clip at a  
23 pseudo-randomly selected position within the range from the  $E_{\text{MIN}}$  to the  $E_{\text{MAX}}$   
24 block and the search space for the detection algorithm is bounded to static time  
25 warping = 10% and DTW dynamic time warping = 6%, the total number of

1 correlation tests performed during the exhaustive search for watermark existence  
2 equals:

$$3 \text{ Tests: } Tests = \frac{E_{\max} - E_{\min}}{M} \frac{2STW}{DTW} \frac{2SFS}{DFS}$$

6 where STW is the static time warp, DTW is the dynamic time warp, SFS is the  
7 static frequency shift, and DFS is the dynamic frequency shift.

8 If the watermark is embedded starting from at earliest the tenth and at the  
9 latest the thirtieth second of the audio clip, this formula indicates that the  
10 exhaustive search would require approximately 17,000 correlation tests. Since  
11 each correlation test requires 153.45 multiply-additions, the computational  
12 complexity of the audio watermarking algorithm for this set of parameters is at the  
13 level of  $10^8$  multiply-additions. Obviously, for a 100MFLOPS machine, the  
14 exhaustive watermark detection process would require approximately one second  
15 of computation time. This performance is realistically expected in real life  
16 applications because all popular Internet music standards MP3 and MSAudio store  
17 the audio content as a compressed collection of frequency magnitude samples.

### 19 Exemplary WMA Implementation

20 Figs. 9 and 10 illustrate the watermark encoding system 32' and watermark  
21 decoding system 52', respectively, integrated into an audio  
22 compression/decompression unit, such as the Windows Media Audio (WMA)  
23 module available from Microsoft Corporation. In Fig. 9, the IMCLT module 80 is  
24 integrated into the WMA encoder 190, which converts the frequency-domain  
25

1 signal  $\{Y_{MAG}(k), \phi(k)\}$  to a time-domain watermarked and encoded signal block  
2  $b(n)$ . In this manner, the compression unit and the watermark encoding system  
3 utilize the same frequency magnitude components for both compression and  
4 watermarking, thereby gaining some computational efficiency. In Fig. 10, the  
5 MCLT module 60 and auditory masking model 62 are integrated into a WMA  
6 decoder 200. Again, this allows the decompression unit (WMA decoder 200) and  
7 the watermark detecting system to utilize the same frequency magnitude  
8 components for both decompression and detection.

### 9 10 Conclusion

11 Although the invention has been described in language specific to structural  
12 features and/or methodological steps, it is to be understood that the invention  
13 defined in the appended claims is not necessarily limited to the specific features or  
14 steps described. Rather, the specific features and steps are disclosed as preferred  
15 forms of implementing the claimed invention.